Alqalam Journal of Medical and Applied Sciences. 2026;9(11):3954-3963
https://doi.org/10.54361/ajmas.261154

Research Article

Brain Tumor Detection and Segmentation Using YOLO11 and SAM2

Mustafa Algaet[1]*, Salima Hindi[2]

[1]*Department of Network, Faculty of Information Technology, Elmergib University, Libya*
[2]*Department of Computer Science, School of Basic Science, the Libyan Academy, Libya*
Email. malgaet@elmergib.edu.ly

Abstract
Accurate delineation of brain tumors in magnetic resonance imaging (MRI) is important for diagnosis, treatment planning, and longitudinal assessment. Manual interpretation and segmentation are time-consuming and may vary between observers. This study evaluates an integrated deep learning framework that combines YOLO11-s for tumor localization with the Segment Anything Model 2 (SAM2) for prompt-guided pixel-level segmentation. The framework uses multimodal MRI information from T1-weighted, T2-weighted, and fluid-attenuated inversion recovery (FLAIR) sequences. It was trained and evaluated on a processed subset of the BraTS 2023 dataset and subsequently assessed using clinical MRI data collected from Zliten Medical Center and Misurata Medical Center in Libya. Five YOLO11 variants were compared under a common training configuration. YOLO11-s provided the strongest overall balance of detection performance and inference time, achieving 96.10% precision, 97.06% recall, an F1-score of 96.58%, and mAP@0.5 of 95.01% on the BraTS 2023 test set. Its inference time was 11.88 ms per image. Bounding boxes predicted by YOLO11-s were passed to SAM2 without task-specific fine-tuning. The integrated framework achieved a mean Dice coefficient of 0.7190 and mean intersection over union (IoU) of 0.6267 on 433 test images. On the external hospital evaluation subset, detection mAP@0.5 was 93.50%, while segmentation achieved a mean Dice of 0.6855 and mean IoU of 0.5718 on 26 images. These results indicate that detector-guided SAM2 segmentation is a promising modular approach, although the decrease in stricter localization and segmentation metrics highlights the need for broader clinical validation.
Keywords. Brain Tumor, Magnetic Resonance Imaging; YOLO11, SAM2, Object Detection, Medical Image.

Introduction

Gliomas, including aggressive forms such as glioblastoma, are particularly challenging because of their heterogeneous appearance, infiltrative boundaries, and variable growth patterns. Timely and reliable imaging assessment is therefore an important component of patient management [6,19]. Magnetic resonance imaging is widely used in brain tumor assessment because it provides high soft-tissue contrast without ionizing radiation. Different MRI sequences reveal complementary properties. T1-weighted images provide anatomical detail; T2-weighted images emphasize fluid-rich regions and edema; and FLAIR suppresses cerebrospinal-fluid signal while making many pathological abnormalities more conspicuous. Contrast-enhanced T1 imaging can also highlight enhancing tumor components. The present study focuses on T1, T2, and FLAIR because these modalities were available in the benchmark and clinical data sources used to establish a consistent processing pipeline [7].

Despite the diagnostic value of MRI, manual tumor localization and delineation remain labor-intensive. Tumors can exhibit irregular shapes, heterogeneous signal intensity, poorly defined boundaries, and substantial variation in size. These factors complicate reproducible segmentation and may increase the time required for image review. Automated systems that can identify a suspicious region and generate a detailed mask may help reduce repetitive annotation work and provide consistent visual support [15].

Deep learning has advanced medical image analysis by learning hierarchical representations directly from image data. Convolutional neural networks have been applied to classification and segmentation, while single-stage object detectors such as You Only Look Once (YOLO) provide rapid localization through bounding-box prediction. However, a detection box does not provide the pixel-level contour needed for detailed volumetric measurement or boundary inspection. Conversely, segmentation systems may require dense annotated training data and may be computationally demanding [17]. This study evaluates a two-stage approach in which YOLO11 detects and localizes tumor regions, and SAM2 uses the predicted bounding boxes as prompts to generate segmentation masks. The detector narrows the region of interest for the prompt-guided segmentation model [14].

Research problem and motivation

The research problem addressed in this study is the need for an integrated workflow that combines efficient tumor localization with pixel-level delineation while retaining the ability to operate across benchmark and clinical MRI data. YOLO11 offers fast object detection, but its bounding boxes are coarse representations of tumor extent. SAM2 can produce detailed masks from spatial prompts, but it requires a prompt that identifies the relevant region. Passing YOLO11 predictions to SAM2 provides a direct interface between these complementary capabilities.

Aim and objectives

The overall aim is to develop and evaluate an integrated YOLO11–SAM2 framework for automated brain tumor detection and segmentation from MRI slices. The specific objectives are to: (i) preprocess multimodal MRI data into a standardized

Copyright *Author (s) 2026. Distributed under Creative Commons CC BY 4.0*
Received*: 19-09-2026 -* ***Accepted****: 18-11-2026 -* ***Published****: 26-11-2026*

image representation; (ii) compare five YOLO11 model sizes for tumor detection; (iii) select a detector based on the balance between accuracy and inference time; (iv) use predicted bounding boxes to prompt SAM2; and (v) evaluate detection and segmentation performance on the BraTS 2023 benchmark and an external clinical dataset.

Research questions

The study is guided by three practical questions. First, how accurately can the integrated framework detect and segment tumor regions in the prepared BraTS 2023 test data? Second, which YOLO11 variant provides the most appropriate balance between detection quality and computational cost? Third, to what extent does the selected detector–segmenter pipeline retain performance when applied to clinical MRI examinations acquired at two medical centers?

Contributions

This study makes four contributions: (i) a multimodal 2D MRI preprocessing workflow using T1, FLAIR, and T2 channels; (ii) a comparison of YOLO11-n, YOLO11-s, YOLO11-m, YOLO11-l, and YOLO11-x under a common training configuration; (iii) integration of the selected YOLO11-s detector with SAM2 using bounding-box prompts, without task-specific SAM2 fine-tuning; and (iv) benchmark and external clinical evaluations.

Scope

The framework operates on two-dimensional axial MRI slices rather than complete three-dimensional volumes. The BraTS 2023 dataset supplies annotated masks for quantitative benchmark evaluation. The hospital data are used as an external clinical evaluation source; manual reference annotations were prepared for the subsets used in the reported quantitative assessment. The results should therefore be interpreted as evidence from the evaluated image subsets, not as proof of clinical diagnostic efficacy or autonomous clinical readiness.

Related work

Traditional brain tumor analysis methods have used handcrafted image descriptors and conventional classifiers, including support vector machines, k-nearest neighbors, decision trees, and logistic regression. These methods depend on manually designed features and may be sensitive to variations in intensity, texture, and acquisition conditions. Reviews of brain tumor segmentation and classification describe the transition from handcrafted pipelines to deep neural networks that learn complex representations from MRI data [4,18].

CNN-based approaches have demonstrated strong performance in tumor classification and segmentation. DeepSeg, for example, adapted U-Net-based architectures for automated segmentation using FLAIR MRI and BraTS data [21]. Other studies have explored optimized U-Net designs, attention mechanisms, and hybrid architectures to improve boundary delineation. These methods are important reference points because they are trained specifically for segmentation and can achieve high overlap scores, although their performance depends on architecture, labels, and evaluation protocol.

YOLO is a single-stage object-detection family that predicts object locations and classes in one forward pass. Its speed and relatively straightforward inference pipeline have encouraged its use in medical image localization. Studies have investigated YOLOv5, YOLOv7, YOLOv8, YOLOv10, and YOLOv11 for brain tumor detection and related tasks [3,12]. YOLO-based systems can also provide regions of interest to downstream segmentation models, creating modular detection–segmentation pipelines.

The Segment Anything Model introduced prompt-based segmentation, while SAM2 extended the framework to image and video segmentation [14]. In medical imaging, SAM and SAM2 have been investigated for tumor delineation, including the use of bounding-box prompts. Ali et al. [2] evaluated SAM on brain tumor MRI, and subsequent studies explored combinations of YOLO detectors and SAM-family models.

Hybrid systems illustrate the value of combining localization and segmentation. YOLO-based detectors have been paired with U-Net or SAM to create integrated analysis pipelines [8,20]. Ratnakar et al. [13] examined YOLOv10 with SAM2.1 for MRI-based neurological analysis. The present study evaluates YOLO11-s as the localization stage and SAM2 as the prompt-guided segmentation stage.

Research gap

A comparatively limited body of work has investigated the integration of YOLO11 with SAM2 for brain tumor detection and segmentation. This study addresses that gap through a unified 2D workflow, comparative detector selection, and evaluation on both BraTS 2023 and local clinical MRI data. The contribution is an empirical integration and cross-dataset assessment rather than a new YOLO11 or SAM2 architecture.

Materials and Methods

Study design and workflow

The study follows an experimental design in which a public benchmark dataset is used to prepare and train the detection component, followed by evaluation of the integrated pipeline on benchmark test images and external clinical MRI images. The processing workflow comprises data acquisition, modality selection, slice extraction, intensity normalization, three-

Copyright *Author (s) 2026. Distributed under Creative Commons CC BY 4.0*
Received*: 19-09-2026 -* ***Accepted****: 18-11-2026 -* ***Published****: 26-11-2026*

channel image construction, mask processing, bounding-box generation, detector training and selection, SAM2 prompting, and quantitative or qualitative evaluation.

The proposed system operates sequentially. First, a preprocessed MRI slice is passed to YOLO11-s, which predicts a tumor bounding box. Second, the predicted box is converted into a spatial prompt for SAM2. Third, SAM2 produces a pixel-level mask for the region of interest. Detection and segmentation outputs are retained for evaluation and visualization. This design assigns tumor localization to the detector and uses the segmentation model to refine the detected region. The workflow is illustrated in Figure 1.

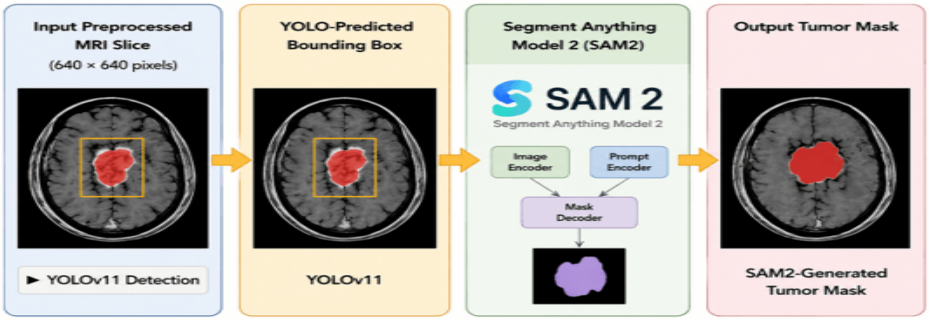


Figure 1. Overall architecture of the proposed YOLO11–SAM2 framework.

Datasets

The primary dataset was BraTS 2023, a benchmark collection of brain MRI volumes stored in NIfTI format with expert-annotated segmentation labels. The dataset includes four conventional modalities: T1, T1CE, T2, and FLAIR. T1, T2, and FLAIR were selected for this pipeline to maintain modality availability across the benchmark and hospital sources. These three sequences were used to construct a composite input image.

The external clinical data were retrospectively collected from Misurata Medical Center and Zliten Medical Center in Libya. The examinations were stored in DICOM format and acquired using different scanners and imaging protocols. Patient-identifying information was removed before analysis. The clinical dataset was used to assess generalization under routine clinical conditions and was not used to train the YOLO11 models.

Dataset composition

After preprocessing and case-wise selection, the BraTS 2023 detection dataset comprised 2,882 two-dimensional axial slices. The dataset composition reported for this study was approximately 85% tumor-positive slices and 15% tumor-negative slices. The split was performed at the patient/case level to reduce data leakage: 70% of cases were assigned to training, 15% to validation, and 15% to testing. The test set contained 433 images for segmentation evaluation. The external clinical segmentation evaluation included 26 images.

MRI preprocessing

BraTS 2023 volumes were loaded from NIfTI files and processed slice by slice. Axial slices were extracted while retaining their slice indices. Slices containing labeled tumor voxels were selected as positive examples, and tumor-negative slices were retained as negative examples. Patient-level partitioning ensured that slices from a given patient were not distributed across multiple subsets. The preprocessing diagram has been revised to show the retention of the reported tumor-negative subset.

Intensity normalization was applied to reduce the influence of extreme voxel values and improve consistency across scans. Intensities were clipped to the 1st and 99th percentiles, scaled to the 0–255 range, and converted to 8-bit images. This procedure produced a consistent intensity representation for detector input.

The three selected MRI modalities were combined into a three-channel composite image. In the OpenCV BGR representation, blue corresponded to T1, green to FLAIR, and red to T2; the channel order was therefore [T1, FLAIR, T2]. This early-fusion representation enabled the model to learn complementary patterns from the three MRI sequences while retaining a three-channel input compatible with the pretrained YOLO architecture. The preprocessing and composite-image construction workflow is summarized in Figure 2.

Tumor masks were processed to generate binary masks and detection annotations. Connected tumor regions were identified, small components were filtered, and bounding boxes were generated around the remaining tumor regions. Composite images and masks were resized to 640 × 640 pixels. The resulting image–annotation pairs were stored in the format required for YOLO training.

Copyright *Author (s) 2026. Distributed under Creative Commons CC BY 4.0*
Received*: 19-09-2026 -* ***Accepted****: 18-11-2026 -* ***Published****: 26-11-2026*

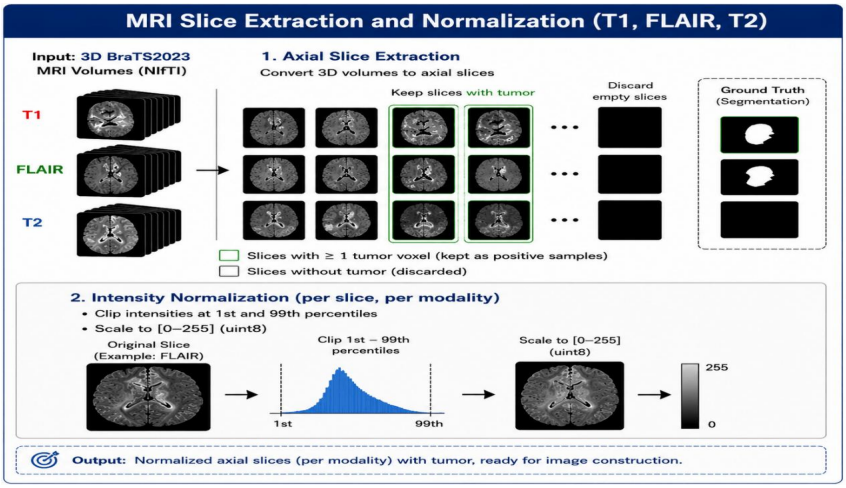


Figure 2. MRI slice extraction, normalization, retention of tumor-positive and tumor-negative slices, and T1–FLAIR–T2 composite construction.

Clinical MRI preprocessing

The clinical DICOM workflow began with series indexing and modality identification, followed by ordering slices according to the acquisition sequence. Available T1, FLAIR, and T2 series were visually reviewed and selected where present. Axial two-dimensional slices were generated, and corresponding modalities were matched when a composite image was required. Robust intensity normalization was used to reduce variation caused by scanner and acquisition settings. Low-information slices containing limited brain tissue were excluded, and the brain region was cropped to reduce irrelevant background. Images were resized to 640 × 640 pixels. Patient-identifying information was removed before clinical-image processing. The institutional ethics approval identifier, approval date, and applicable consent or waiver information were not available in the manuscript materials reviewed for this revision. These details must be inserted from the approved institutional documentation before submission.

YOLO11 tumor detection

YOLO11 served as the object-detection stage of the proposed framework. The detector predicted bounding boxes around tumor regions in the three-channel MRI composites. Training annotations were generated from the BraTS segmentation masks by calculating the minimum enclosing rectangle around each tumor region, thereby converting pixel-level annotations into the bounding-box format required for object detection.

Five model variants were evaluated: YOLO11-n, YOLO11-s, YOLO11-m, YOLO11-l, and YOLO11-x. All variants used the same dataset, preprocessing pipeline, and input resolution of 640 × 640 pixels. The Ultralytics YOLO11 training framework and its augmentation procedures were used, and the checkpoint with the best validation performance was retained. The complete configuration for each model must be reported to permit exact replication.

The available experimental record documents 100 training epochs for YOLO11-s. The epoch counts used for YOLO11-n, YOLO11-m, YOLO11-l, and YOLO11-x, as well as the batch size, optimizer, learning rate, weight decay, hardware model and GPU memory, software versions, and random seeds, were not specified in the manuscript materials. These values must be recovered from the original training notebooks or experiment logs and added to the final Methods section; they cannot be inferred reliably from the reported performance metrics.

Selection of the detection model

Model selection considered precision, recall, F1-score, mAP@0.5, mAP@0.5:0.95, and inference time. The objective was not simply to select the largest model or the fastest model, but to identify a suitable compromise between localization performance and computational cost. YOLO11-s was selected because it achieved the highest reported precision, recall, F1-score, and mAP values among the five evaluated variants while retaining a low inference time.

SAM2-based segmentation

The selected YOLO11-s model provides the bounding-box prompts for SAM2. For each detected tumor, the predicted box is passed to SAM2 as a spatial prompt, and SAM2 generates a segmentation mask within the corresponding region. SAM2 was applied in inference mode without task-specific fine-tuning. The resulting mask is used for visual inspection and, where reference masks are available, quantitative evaluation. This is an important characteristic of the study: the segmentation component is used as a pretrained promptable model rather than retrained on the study's MRI annotations.

Integrated inference procedure

For each input image, the integrated inference procedure consists of the following steps. (1) Load and preprocess the MRI slice. (2) Construct the three-channel T1–FLAIR–T2 composite. (3) Resize the image to 640 × 640 pixels. (4) Run YOLO11-s to obtain tumor bounding boxes and confidence scores. (5) Convert each predicted bounding box into a SAM2 prompt. (6)

Copyright *Author (s) 2026. Distributed under Creative Commons CC BY 4.0*
Received*: 19-09-2026 -* ***Accepted****: 18-11-2026 -* ***Published****: 26-11-2026*

Run SAM2 to generate the pixel-level segmentation mask. (7) Save or display the detection box and segmentation overlay. The same overall sequence is used for benchmark and clinical images.

The detection and segmentation stages have a dependency: SAM2 receives the predicted box rather than an independently supplied ground-truth box. Consequently, errors in localization can propagate into the segmentation stage. A box that excludes part of a tumor may constrain SAM2 away from the omitted region, while an excessively large box may include surrounding tissue and complicate mask generation. This dependency is considered when interpreting the results.

Evaluation metrics

Detection performance was measured using precision, recall, F1-score, mean average precision at an intersection-over-union threshold of 0.5 (mAP@0.5), and mAP averaged over IoU thresholds from 0.50 to 0.95 (mAP@0.5:0.95). Precision measures the proportion of predicted positive detections that are correct; recall measures the proportion of reference tumor objects that are detected. The F1-score is the harmonic mean of precision and recall. mAP summarizes average precision across classes and IoU criteria.

Segmentation performance was evaluated using the Dice similarity coefficient and intersection over union (IoU). For predicted mask P and reference mask G, Dice is calculated as $2|P \cap G|/(|P| + |G|)$, whereas IoU is $|P \cap G|/|P \cup G|$. Both metrics range from 0 to 1, with larger values indicating greater overlap between predicted and reference regions. Mean Dice and mean IoU were calculated across the evaluated images.

On the BraTS 2023 test set, detection and segmentation metrics were calculated against the available reference annotations. For the external hospital data, reference bounding boxes and masks were manually created using the Computer Vision Annotation Tool (CVAT) for the quantitative evaluation subsets. The reported hospital segmentation metrics therefore apply to 26 annotated images, rather than to every clinical examination collected. The professional qualifications of the annotators, the number of annotators, the annotation instructions, and any inter-annotator agreement assessment were not documented in the materials available for this revision and must be supplied by the authors.

Implementation summary

The framework was implemented with Python and GPU-enabled deep learning libraries in Google Colab. NiBabel supported NIfTI reading, pydicom supported DICOM processing, and OpenCV and NumPy supported image normalization and composite-image generation. Ultralytics YOLO11 performed detection, while SAM2 performed prompt-guided segmentation. Pandas and Matplotlib supported result organization and visualization.

Experimental results *Comparison of YOLO11 variants on BraTS 2023*

All five YOLO11 variants achieved high detection performance on the BraTS 2023 test set. Precision and recall exceeded 95% for every model. Differences were more apparent in mAP and inference time. YOLO11-n was the fastest variant at 10.79 ms per image, but its detection scores were slightly below those of YOLO11-s. Larger variants required substantially more time without providing better overall detection performance. The detection results for the five YOLO11 variants are summarized in (Table 1).

Table 1. Performance comparison of YOLO11 variants on the BraTS 2023 test set

Model	Precision	Recall	mAP@0.5	mAP@0.5:0.95	F1-score	Inference (ms/image)
YOLO11-n	0.9608	0.9651	0.9475	0.7916	0.9629	10.79
YOLO11-s	0.9610	0.9706	0.9501	0.8138	0.9658	11.88
YOLO11-m	0.9601	0.9633	0.9395	0.8002	0.9617	22.22
YOLO11-l	0.9555	0.9577	0.9378	0.8099	0.9566	28.11
YOLO11-x	0.9538	0.9561	0.9397	0.7954	0.9550	57.52

YOLO11-s achieved the highest precision (96.10%), recall (97.06%), F1-score (96.58%), mAP@0.5 (95.01%), and mAP@0.5:0.95 (81.38%). Its inference time was 11.88 ms per image, only slightly slower than YOLO11-n. The larger YOLO11-m, YOLO11-l, and YOLO11-x variants did not improve the reported detection metrics and incurred higher inference times. YOLO11-s was therefore selected for integration with SAM2.

Detection performance on the BraTS 2023 test set

The selected YOLO11-s model achieved precision of 0.9610 and recall of 0.9706. The corresponding F1-score was 0.9658. The mAP@0.5 value of 0.9501 indicates strong detection performance at the 0.5 IoU threshold, while mAP@0.5:0.95 of 0.8138 provides a more stringent assessment of localization quality across multiple thresholds. The evaluation also includes precision–recall curves, a confusion matrix, and representative detections to illustrate model behavior.

Segmentation performance on BraTS 2023

The bounding boxes predicted by YOLO11-s were used as prompts for SAM2. The integrated model achieved a mean Dice coefficient of 0.7190 and a mean IoU of 0.6267 on 433 test images. These values indicate meaningful overlap between the predicted tumor masks and the expert reference masks, while also showing that segmentation remains less accurate than

Copyright *Author (s) 2026. Distributed under Creative Commons CC BY 4.0*
Received*: 19-09-2026 -* ***Accepted****: 18-11-2026 -* ***Published****: 26-11-2026*

detection localization metrics might suggest. Detection and segmentation metrics measure different tasks and should not be interpreted as directly interchangeable. Representative segmentation outputs from the integrated framework are shown in (Figure 3).

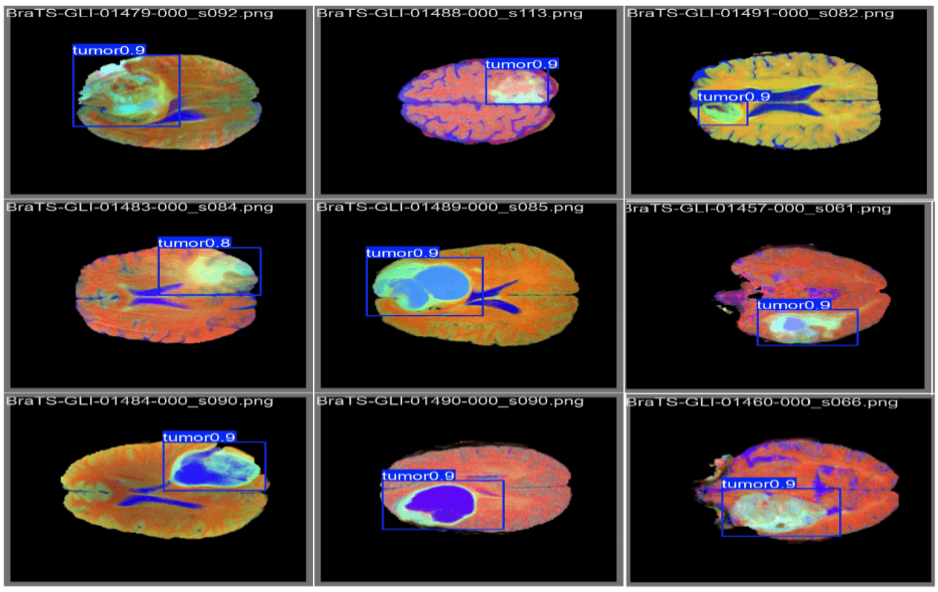


Figure 3. YOLO11-s detection examples showing predicted bounding boxes. SAM2 segmentation masks and ground-truth contours are not displayed in the currently available panels.

The quantitative segmentation results on the BraTS 2023 test set are presented in (Table 2).

Table 2. Segmentation performance of the YOLO11-s–SAM2 framework on the BraTS 2023 test set.

Segmentation metric	BraTS 2023 test result
Mean Dice coefficient	0.7190
Mean IoU	0.6267
Number of test images	433

SAM2 was used in inference mode without task-specific fine-tuning. This demonstrates the use of a pretrained promptable segmentation model with automatically generated bounding-box prompts. The mean Dice score also indicates scope for improvement, particularly around irregular tumor boundaries and regions with weak contrast against adjacent tissue.

External hospital detection results

The framework was next evaluated on clinical MRI images collected from Zliten Medical Center and Misurata Medical Center. Because the hospital dataset did not originally contain detection annotations, reference bounding boxes were manually created using CVAT for the reported quantitative assessment. YOLO11-s achieved precision of 0.9089, recall of 0.9167, and an F1-score of 0.9128. Its mAP@0.5 was 0.9350, while mAP@0.5:0.95 was 0.4401. The external hospital detection results are summarized in (Table 3).

Table 3. External hospital detection performance of YOLO11-s.

Detection metric	External hospital result
Precision	0.9089 (90.89%)
Recall	0.9167 (91.67%)
F1-score	0.9128 (91.28%)
mAP@0.5	0.9350 (93.50%)
mAP@0.5:0.95	0.4401 (44.01%)

The difference between mAP@0.5 and mAP@0.5:0.95 is substantial. This suggests that detections frequently overlap the reference tumor region sufficiently at the 0.5 threshold, but precise box alignment becomes more difficult under stricter IoU requirements. Variation in image quality, tumor characteristics, and acquisition protocols may contribute to this reduction.

External hospital segmentation results

The predicted hospital-image bounding boxes were passed to SAM2 to generate segmentation masks. Since the original clinical dataset did not include ground-truth segmentation masks, reference masks were manually prepared with CVAT for

Copyright *Author (s) 2026. Distributed under Creative Commons CC BY 4.0*
Received*: 19-09-2026 -* ***Accepted****: 18-11-2026 -* ***Published****: 26-11-2026*

the subset used in quantitative evaluation. Across 26 evaluated images, the integrated framework achieved a mean Dice coefficient of 0.6855 and a mean IoU of 0.5718. The external hospital segmentation results for the 26 evaluated images are reported in (Table 4).

Table 4. External hospital segmentation performance of the YOLO11-s–SAM2 framework

Segmentation metric	External hospital result
Mean Dice coefficient	0.6855
Mean IoU	0.5718
Number of evaluated images	26

The external results are lower than those obtained on BraTS 2023, but the decline is moderate for Dice and IoU. The findings suggest that the integrated pipeline retained useful segmentation performance on the evaluated clinical images despite differences in acquisition conditions. However, the hospital sample used for quantitative segmentation is small, and the results should be treated as preliminary external validation.

Cross-dataset comparison

A comparison of the benchmark and external hospital results is provided in (Table 5).

Table 5. Cross-dataset comparison of detection and segmentation performance

Metric	BraTS 2023	Hospital data	Difference (hospital – BraTS)
Precision (%)	96.10	90.89	−5.21
Recall (%)	97.06	91.67	−5.39
F1-score (%)	96.58	91.28	−5.30
mAP@0.5 (%)	95.01	93.50	−1.51
mAP@0.5:0.95 (%)	81.38	44.01	−37.37
Dice (%)	71.90	68.55	−3.35
IoU (%)	62.67	57.18	−5.49

The cross-dataset comparison shows that mAP@0.5 decreased by only 1.51 percentage points, whereas mAP@0.5:0.95 decreased by 37.37 percentage points. Dice decreased by 3.35 percentage points and IoU by 5.49 percentage points. The largest change is therefore associated with strict bounding-box localization, which appears more sensitive to differences between the benchmark and clinical imaging sources than the mAP@0.5 result.

These comparisons should be interpreted cautiously because the two datasets differ in source, image acquisition, sample composition, and annotation procedures. The hospital segmentation result is based on 26 manually annotated images, whereas the BraTS test result is based on 433 images with benchmark reference masks.

Comparison with selected related methods

The comparison with selected published methods is contextual because the studies use different datasets, dimensionalities, tasks, and evaluation protocols. Their numerical results should not be interpreted as a controlled head-to-head comparison. Table 6 summarizes the methods and results reported in the corresponding publications.

Table 6. Comparison of the proposed framework with selected related methods

Method	Dataset / type	Task	Reported result in the cited study
YOLOv10-X (Chhimpa et al., 2026) [5]	Br35H; external evaluation on BraTS 2021 2D slices	Detection	BraTS 2021: precision 0.942; recall 0.915; mAP@0.5 0.953
Advanced nnU-Net (Kharaji et al., 2024) [9]	BraTS glioma and pediatric cohorts; 3D	Segmentation	Mean Dice: 0.830 (adult glioma); 0.710 (pediatric cohort)
SAM2-UNet (Pidishetti et al., 2025) [11]	BraTS 2020; 2D	Segmentation	Mean Dice 0.771; mean IoU 0.569
YOLOv5 + 2D U-Net (Ahsan et al., 2025) [1]	Brain Tumor Figshare and BraTS 2018; 2D	Detection + segmentation	mAP 0.895; Dice 0.881
YOLOv8 + SAM ViT-B (Kassam et al., 2024) [8]	BraTS 2021 with noise-augmented images; 2D	Detection + segmentation	Dice 0.790 for enhancing tumor
Proposed YOLO11-s + SAM2	BraTS 2023; 2D	Detection + segmentation	Precision 0.961; recall 0.971; mAP@0.5 0.950; Dice 0.719; IoU 0.627

The proposed framework provides an integrated detection–segmentation workflow and reports strong detection performance. Its Dice score is lower than values reported for some dedicated segmentation networks and hybrid systems. This difference may reflect the use of SAM2 without task-specific fine-tuning, whereas some comparator methods were

Copyright *Author (s) 2026. Distributed under Creative Commons CC BY 4.0*
Received*: 19-09-2026 -* ***Accepted****: 18-11-2026 -* ***Published****: 26-11-2026*

trained directly for segmentation. Differences in datasets and evaluation protocols prevent a definitive ranking across studies. In addition, no independent segmentation baseline was trained and evaluated on the same 433-image BraTS 2023 test split in the experiments reported here; a controlled comparison requires an additional experiment using the identical split and evaluation procedure.

Summary of main findings

Three findings emerge. First, YOLO11-s was the strongest overall detector among the five tested variants, combining high detection metrics with an inference time of 11.88 ms per image. Second, SAM2 generated useful masks from detector-derived prompts without fine-tuning, achieving a mean Dice of 0.7190 on BraTS 2023. Third, the integrated pipeline retained encouraging performance on the external clinical subset, although strict localization and segmentation metrics declined. Together, these findings support the feasibility of the proposed modular design while identifying clear opportunities for improvement.

Discussion

This study evaluated a sequential detection–segmentation framework for brain tumor MRI analysis. The principal design choice was to use YOLO11-s to identify the tumor region and SAM2 to refine the localization into a pixel-level mask. The results support the usefulness of this division of tasks: YOLO11-s provided high detection precision and recall, and SAM2 generated segmentation masks without being fine-tuned on the study dataset.

The comparison of YOLO11 variants shows that increasing model size did not improve detection performance in this experiment. YOLO11-s outperformed the other tested variants on precision, recall, F1-score, mAP@0.5, and mAP@0.5:0.95, while requiring substantially less inference time than the medium, large, and extra-large variants. YOLO11-n was slightly faster, but its detection metrics were lower. This supports the selection of YOLO11-s as a practical compromise for the integrated workflow.

The segmentation results require a more nuanced interpretation. A mean Dice of 0.7190 on BraTS 2023 indicates useful but incomplete overlap with reference tumor masks. SAM2 was not fine-tuned, and its segmentation quality depended on the quality of the YOLO11-s bounding-box prompt. This modular approach reduces the need for training a separate segmentation network, but it also creates an error-propagation pathway: a missed detection or inaccurate box can constrain the mask prediction.

The external evaluation provides an initial indication of generalization beyond the public benchmark. Detection mAP@0.5 remained at 93.50%, and the mean Dice was 0.6855 on the 26-image clinical segmentation subset. Nevertheless, the reduction in mAP@0.5:0.95 indicates that accurate boundary localization is more difficult under clinical acquisition variability. This may reflect differences in scanner characteristics, image quality, tumor appearance, and annotation conditions. Further multi-center validation is needed before making stronger claims about generalization.

Practical implications

A detector-guided segmentation pipeline may help organize MRI review by producing a candidate tumor location and a corresponding mask for inspection. It may reduce some manual interaction required to define a region of interest. The current results, however, do not establish clinical diagnostic accuracy, patient outcome benefit, or suitability for unsupervised use. Outputs should be regarded as research results requiring expert review.

Limitations

Several limitations should be considered when interpreting the findings. First, the framework processes 2D axial slices rather than complete 3D MRI volumes. Spatial continuity between adjacent slices is therefore not explicitly modeled, and the output does not directly represent a volumetric tumor contour. Second, the hospital segmentation evaluation includes only 26 images, which is a limited sample for assessing performance across clinical populations, scanners, and acquisition protocols.

Third, SAM2 was used without task-specific fine-tuning, which may limit adaptation to MRI-specific tumor boundaries and signal characteristics. Fourth, segmentation depends on the quality of the YOLO11-s bounding boxes, and detection errors can propagate into the segmentation stage. Fifth, the manuscript materials do not contain the complete training configuration or hardware specifications, limiting exact reproducibility until the original experiment logs are recovered.

Sixth, the clinical evaluation required manual annotation. The reported metrics depend on the annotation protocol and selected subset; however, annotator qualifications and inter-annotator agreement were not documented in the available materials. Finally, the study does not establish prospective clinical utility, workflow impact, or diagnostic performance relative to radiologists.

Future research

Future work should investigate fine-tuning SAM2 or adapting its prompt encoder and mask decoder using curated medical MRI data. The effect of different prompt strategies, including bounding boxes, points, and combined prompts, should be measured systematically. Extending the system to 3D MRI volumes could exploit inter-slice context and improve volumetric

Copyright *Author (s) 2026. Distributed under Creative Commons CC BY 4.0*
Received*: 19-09-2026 -* ***Accepted****: 18-11-2026 -* ***Published****: 26-11-2026*

consistency. Larger multi-center datasets acquired with different scanners and protocols are also needed to evaluate robustness.

Further studies should report complete implementation details, including hardware, software versions, batch size, optimizer, learning rate, training duration, random seeds, and model checkpoint selection. Clinical validation should include clearly documented ethics approval, data governance, annotation procedures, and inter-rater agreement. A user-facing prototype could then be evaluated as a radiologist-assistance tool, with the clinician retaining responsibility for interpretation and decisions.

Conclusion

This study presents an integrated deep learning framework for brain tumor detection and segmentation from MRI images using YOLO11-s and SAM2. The workflow combines multimodal MRI preprocessing, single-stage tumor detection, and bounding-box-guided prompt segmentation. T1, FLAIR, and T2 sequences were fused into three-channel images, and five YOLO11 variants were compared under a common training configuration.

YOLO11-s achieved the strongest overall detection performance among the evaluated variants, with 96.10% precision, 97.06% recall, 96.58% F1-score, 95.01% mAP@0.5, and 81.38% mAP@0.5:0.95. Its inference time was 11.88 ms per image. When its predicted boxes were passed to SAM2 without task-specific fine-tuning, the integrated framework achieved a mean Dice of 0.7190 and mean IoU of 0.6267 on 433 BraTS 2023 test images.

On the external clinical evaluation subset from Zliten and Misurata Medical Centers, the detector achieved 93.50% mAP@0.5, and the integrated segmentation pipeline achieved a mean Dice of 0.6855 and mean IoU of 0.5718 on 26 images. These results suggest that the proposed framework can retain useful performance across different MRI sources, while the decline in strict localization and segmentation measures indicates that further adaptation and validation are necessary.

Overall, the study supports YOLO11-s–SAM2 as a flexible research framework for automated tumor localization and mask generation. The current evidence is promising but preliminary. Larger clinical datasets, 3D processing, SAM2 fine-tuning, complete reproducibility reporting, and prospective evaluation are required before the system can be considered for routine clinical deployment.

Acknowledgments

The authors would like to express their sincere appreciation to the Department of Network, Faculty of Information Technology, Elmergib University, and the Department of Computer Science, School of Basic Science, the Libyan Academy, for their academic and institutional support. The authors also gratefully acknowledge Zliten Medical Center and Misurata Medical Center for their cooperation and for providing access to the clinical MRI data used in this study.

Conflicts of Interest

The authors declare no conflicts of interest. No personal, professional, or financial relationships influenced the design of the study, data collection, analysis, interpretation of results, or preparation of the manuscript.

References

1. Ahsan R, Shahzadi I, Najeeb F, Omer H. Brain tumor detection and segmentation using deep learning. MAGMA. 2025;38(1):13–22. doi:10.1007/s10334-024-01203-5.
2. Ali L, Alnajjar F, Swavaf M, Elharrouss O, Abd-alrazaq A, Damseh R. Evaluating segment anything model (SAM) on MRI scans of brain tumors. Sci Rep. 2024;14. doi:10.1038/s41598-024-72342-x.
3. Almufareh MF, Imran M, Khan A, Humayun M, Asim M. Automated brain tumor segmentation and classification in MRI using YOLO-based deep learning. IEEE Access. 2024;12:16189–207. doi:10.1109/ACCESS.2024.3359418.
4. Biratu ES, Schwenker F, Ayano YM, Debelee TG. A survey of brain tumor segmentation and classification algorithms. J Imaging. 2021;7(9). doi:10.3390/jimaging7090179.
5. Chhimpa GR, Awasthi S, Bhati N, Yadav P, Wani NA. A transfer learning-driven fine-tuning of YOLOv10 for improved brain tumor detection in MRI images. Sci Rep. 2026;16:98. doi:10.1038/s41598-025-28813-w.
6. Fahim TA, Alam FB, Hossain MA. Brain tumor detection, classification and segmentation by deep learning models from MRI images: Recent approaches, challenges and future directions. Array. 2025;28:100571. doi:10.1016/j.array.2025.100571.
7. Hamim SA, Jony AI. Enhancing brain tumor MRI segmentation accuracy and efficiency with optimized U-Net architecture. Malays J Sci Adv Technol. 2024;4(3):197–202. doi:10.56532/mjsat.v4i3.302.
8. Kassam S, Markham A, Vo K, Revanakara Y, Lam M, Zhu K. Intraoperative glioma segmentation with YOLO + SAM for improved accuracy in tumor resection [Preprint]. arXiv:2408.14847. 2024.
9. Kharaji M, Abbasi H, Orouskhani Y, Shomalzadeh M, Kazemi F, Orouskhani M. Brain tumor segmentation with advanced nnU-Net: Pediatrics and adults tumors. Neurosci Inform. 2024;4(2):100156. doi:10.1016/j.neuri.2024.100156.
10. Mostafa AM, Zakariah M, Aldakheel EA. Brain tumor segmentation using deep learning on MRI images. Diagnostics. 2023;13(9). doi:10.3390/diagnostics13091562.
11. Pidishetti RV, Amjad M, Sheng VS. Advanced brain tumor segmentation using SAM2-UNet. Appl Sci. 2025;15(6):3267. doi:10.3390/app15063267.
12. Priyadharshini S, Bhoopalan R, Manikandan D, Ramaswamy K. A successive framework for brain tumor interpretation using YOLO variants. Sci Rep. 2025;15. doi:10.1038/s41598-025-13155-4.

Copyright *Author (s) 2026. Distributed under Creative Commons CC BY 4.0*
Received*: 19-09-2026 -* ***Accepted****: 18-11-2026 -* ***Published****: 26-11-2026*

13. Ratnakar A, Sawant S, Karajagikar J. YOLOv10 and SAM 2.1 for enhanced MRI segmentation and improved neurological disease diagnosis. Int J Adv Res. 2025;13(2):454–83. doi:10.21474/IJAR01/20398.
14. Ravi N, Gabeur V, Hu YT, Hu R, Ryali C, Ma T, et al. SAM 2: Segment anything in images and videos [Preprint]. arXiv:2408.00714. 2024.
15. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27–30; Las Vegas, NV. IEEE; 2016. p. 779–88. doi:10.1109/CVPR.2016.91.
16. Saeedi S, Rezayi S, Keshavarz H, Niakan Kalhori SR. MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. BMC Med Inform Decis Mak. 2023;23. doi:10.1186/s12911-023-02114-6.
17. Sapkota R, Karkee M. Ultralytics YOLO evolution: An overview of YOLO26, YOLO11, YOLOv8 and YOLOv5 object detectors for computer vision and pattern recognition [Preprint]. arXiv:2510.09653. 2026.
18. Tiwari A, Srivastava S, Pant M. Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019. Pattern Recognit Lett. 2020;131:244–60. doi:10.1016/j.patrec.2019.11.020.
19. Ye J. Neoplasia: A comprehensive overview of abnormal cell growth and tumor development. J Carcinog Mutagen. 2024;15:451. doi:10.35248/2157-2518.24.15.451.
20. Zafar W, Husnain G, Iqbal A, Alzahrani AS, Irfan MA, Ghadi YY, et al. Enhanced TumorNet: Leveraging YOLOv8s and U-Net for superior brain tumor detection and segmentation utilizing MRI scans. Results Eng. 2024;24:102994. doi:10.1016/j.rineng.2024.102994.
21. Zeineldin RA, Karar ME, Coburger J, Wirtz CR, Burgert O. DeepSeg: Deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images. Int J Comput Assist Radiol Surg. 2020;15(6):909–20. doi:10.1007/s11548-020-02186-z.

Copyright *Author (s) 2026. Distributed under Creative Commons CC BY 4.0*
Received*: 19-09-2026 -* ***Accepted****: 18-11-2026 -* ***Published****: 26-11-2026*